IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

First Named Inventor :	Michael Seltzer et al.	Appeal No.
Appln. No.:	10/647,586	
Filed :	August 25, 2003	Group Art Unit: 2626
For :	METHOD AND APPARATUS USING HARMONIC-MODEL-BASED FRONT END FOR ROBUST SPEECH RECOGNITION	Examiner: Shah, Paras D.
Docket No.:	M61.12-0550	

BRIEF FOR APPELLANTS

Mail Stop Appeal Brief-Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

**Electronically Filed on
October 28, 2008**

Sir:

This is an Appeal of the Office Action of June 17, 2008 in which claims 1-3, 5-13, 15-17, 20, 24 and 25 were rejected.

REAL PARTY IN INTEREST

Microsoft Corporation, a corporation organized under the laws of the state of Washington, and having offices at One Microsoft Way, Redmond, Washington 98052, has acquired the entire right, title and interest in and to the invention, the application, and any and all patents to be obtained therefor, as set forth in the Assignment recorded on Reel 014432, Frame 0319.

RELATED APPEALS AND INTERFERENCES

There are no known related appeals or interferences that will directly affect or be directly affected by or have a bearing on the Board's decision in this appeal.

STATUS OF THE CLAIMS

I.   Total number of claims in the application.

          Claims in the application are: 1-25

II.   Status of all the claims.

A.	Claims cancelled:		4, 14, 18, 19, 21-23
B.	Claims withdrawn but not cancelled:		
C.	Claims pending:		1-3, 5-13, 15-17, 20, 24, 25
D.	Claims allowed:		---
E.	Claims rejected:		1-3, 5-13, 15-17, 20, 24, 25
F.	Claims Objected to:		---

III.   Claims on appeal

          The claims on appeal are: 1-3, 5-13, 15-17, 20, 24, 25

STATUS OF AMENDMENTS

No amendments were filed after the Final Office Action.

SUMMARY OF THE CLAIMED SUBJECT MATTER

Independent claim 1 provides a method (Fig. 3) of identifying an estimate for a noise-reduced value representing a portion of a noise-reduced speech signal. The method includes decomposing each frame of a noisy speech signal into a harmonic component for the frame and a random component for the frame. (Step 304, Fig. 3, page 13, line 29 - page 15, line 25) For each frame, a separate scaling parameter is determine for at least the harmonic component, wherein determining a scaling parameter for each frame of a harmonic component comprises determining a ratio of an energy of the harmonic component in the frame without the random component of the frame to an energy of the frame of the noisy speech signal. (Step 306, Fig. 3, page 15, line 26 - page 16, line 17) For each frame, the harmonic component of the frame is multiplied by the scaling parameter of the frame for the harmonic component to form a scaled harmonic component for the frame. (Step 310, Fig. 3, page 17, line 13 - page 18, line 3) For each frame, the random component of

the frame is multiplied by a fixed scaling parameter for the random component that is the same for all frames and that is less than 1. (Step 310, Fig. 3, page 17, line 13 - page 18, line 3) This forms a scaled random component for the frame. For each frame, the scaled harmonic component for the frame is summed with the scaled random component for the frame to form the noise-reduced value representing a frame of a noise-reduced speech signal wherein the frame of the noise-reduced speech signal has reduced noise relative to the frame of the noisy speech signal. (Step 310, Fig. 3, page 17, line 13 - page 18, line 3)

Independent claim 13 provides a computer-readable storage medium having computer-executable instructions for performing a series of steps. Those steps include identifying a harmonic component and a random component in a noisy speech signal (Step 304, Fig. 3), wherein identifying the harmonic component comprises modeling the harmonic component as a sum of harmonic sinusoids, each sinusoid having an amplitude parameter. (Page 14, lines 7-15) A weighted sum is formed to produce a noise-reduced value representing a noise-reduced speech signal that has reduced noise compared to the noisy speech signal. (Step 310, Fig. 3, page 17, line 13 - page 18, line 3) The weighted sum is formed by multiplying the harmonic component by a scaling value for the harmonic component to form a scaled harmonic component, multiplying the random component by a scaling value for the random component to form a scaled random component, and adding the scaled harmonic component to the scaled random component to form the noise-reduced value. (Step 310, Fig. 3, page 17, line 13 - page 18, line 3) The scaling value for the harmonic component is different than the scaling value for the random component. (Step 310, Fig. 3, page 17, line 13 - page 18, line 3) In addition, the scaling value for the harmonic component is separately determined for each frame of the noisy speech signal and the scaling value for the random component is fixed for all frames of the noisy speech signal so that the same scaling parameter for the random component is used at each frame of the noisy speech signal. (Step 310, Fig. 3, page 17, line 13 - page 18, line 3)

GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

Claims 1-3, 5, 6, 11-13, 15, 16, 20, and 24 were rejected under 35 U.S.C. § 103(a) as being unpatentable over Laroche et al. ("HNM: A Simple Efficient Harmonic and Noise Model for

Speech" 1993, hereinafter Laroche) in view of Rao Gadde et al. (U.S. Patent No. 7,120,580, hereinafter Rao Gadde) in view of Gao (U.S. Patent Publication 2002/0035470) and in view of Rezayee ("An Adaptive KLT Approach for Speech Enhancement", IEEE Transactions on Speech and Audio Processing, 2001).

Claims 7-10, 17, and 25 were rejected under 35 U.S.C. § 103(a) as being unpatentable over Laroche in view of Rao Gadde in view of Gao in view of Rezayee and in further view of Seltzer (SPHINX III Signal Processing Front End Specification, CMU Speech Group, 1999).

ARGUMENT

§103(a) Rejection of claims 1-3, 5, 6, 11-13, 15, 16, 20, and 24

Claims 1-3, 5, 6, 11-13, 15, 16, 20, and 24 were rejected under 35 U.S.C. § 103(a) as being unpatentable over Laroche et al. ("HNM: A Simple Efficient Harmonic and Noise Model for Speech" 1993, hereinafter Laroche) in view of Rao Gadde et al. (U.S. Patent No. 7,120,580, hereinafter Rao Gadde) in view of Gao (U.S. Patent Publication 2002/0035470) and in view of Rezayee ("An Adaptive KLT Approach for Speech Enhancement", IEEE Transactions on Speech and Audio Processing, 2001).

"Section 103 forbids issuance of a patent when the 'differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which the subject matter pertains.' " KSR Int 'l Co. v. TelefZex, Inc., 127 S. Ct. 1727, 1734 (2007). KSR reaffirms the analytical framework set out in Graham v. John Deere Co. of Kansas City, 383 U.S. 1 (1966), which mandates that an objective obviousness analysis includes: (1) determining the scope and content of the prior art; (2) ascertaining the differences between the prior art and the claims at issue; and (3) resolving the level of ordinary skill in the pertinent art. KSR, 127 S. Ct. at 1734.

Claims 1-3, 6, 11, 12

Claim 1 provides a method of identifying an estimate for a noise-reduced value representing a portion of a noise-reduced speech signal. The method includes decomposing each

frame of a noisy speech signal into a harmonic component for the frame and a random component for the frame. For each frame, a separate scaling parameter is determine for at least the harmonic component, wherein determining a scaling parameter for each frame of a harmonic component comprises determining a ratio of an energy of the harmonic component in the frame without the random component of the frame to an energy of the frame of the noisy speech signal. For each frame, the harmonic component of the frame is multiplied by the scaling parameter of the frame for the harmonic component to form a scaled harmonic component for the frame. For each frame, the random component of the frame is multiplied by a fixed scaling parameter for the random component that is the same for all frames and that is less than 1. This forms a scaled random component for the frame. For each frame, the scaled harmonic component for the frame is summed with the scaled random component for the frame to form the noise-reduced value representing a frame of a noise-reduced speech signal wherein the frame of the noise-reduced speech signal has reduced noise relative to the frame of the noisy speech signal.

Claim 1 is not shown or suggested in the combination of cited art. In particular, none of the references shows or suggests determining a scaling parameter for each frame of a harmonic component by determining the ratio of an energy of the harmonic component in the frame without the random component of the frame to an energy of the frame of the noisy speech signal. In addition, none of the cited references shows or suggests determining a scaling parameter for a harmonic component that changes as the ratio of the energy of the harmonic component to the energy of the noisy speech signal changes while using a fixed scaling parameter for a random component, wherein the fixed scaling parameter is the same for all frames.

I. The Cited Art Does Not Show a Scaling Parameter for a Harmonic Component

In the Final Office Action, Fig. 4 and lines 15-20 of Rao Gadde were cited as showing a step of determining a scaling parameter for a frame of a harmonic component that is multiplied by the harmonic component of the frame and Table III of Rezayee was cited as showing a scaling parameter that is a ratio of an energy of a harmonic component of a frame without a random component of the frame to an energy of a noisy speech signal for the frame. Further, it was asserted that it would be obvious to replace the scaling parameter in Rao Gadde with the ratio in Rezayee. Applicants

respectfully dispute that Rao Gadde shows a scaling parameter for a frame of a harmonic component that is multiplied by the harmonic component of the frame, respectfully dispute that it would be obvious to replace the weights shown in Rao Gadde with the ratio shown in Table III of Rezayee and respectfully dispute that the ratio in Rezayee is a ratio of an energy of a harmonic component in a frame without the random component of the frame to an energy of the frame of the noisy speech signal.

Rao Gadde Does Not Show a Scaling Parameter for a Harmonic Component of a Frame

In Rao Gadde, Guassian Mixture Models of clean speech are formed from clean speech training signals. (Rao Gadde, column 4, lines 51-55). Each Guassian model has a mean M and a covariance C. (Rao Gadde, column 4, lines 57-58). Rao Gadde modifies these clean speech models to form modified models that better reflect noisy speech. (Rao Gadde, column 4, lines 51-55). These modifications are made by altering the means of each Gaussian using Equation 1 in column 4. As shown in Equation 1, the mean for the noisy speech Gaussian is formed by taking a weighted sum of a mean for a clean speech Gaussian and a mean for a noise model, where the mean for the noise model is determined from past low energy frames.

In the Final Office Action, it was asserted that the weight (W) in Equation 1 was the harmonic component scaling parameter that was multiplied by the harmonic component of the frame. However, Rao Gadde does not show W being multiplied by a harmonic component of a frame. Instead, W is multiplied by a mean M of a Gaussian model of clean speech. A mean M of a Gaussian model of clean training speech cannot be read to be a harmonic component of a frame under claim 1, because claim 1 requires that the harmonic component of the frame be decomposed from a frame of a noisy speech signal. The mean M is not decomposed from a frame of a noisy speech signal but instead is a statistical value computed from a plurality of clean speech frames. (Rao Gadde, column 4, lines 51-52) As such, Rao Gadde does not show determining a scaling parameter for a harmonic component or multiplying a scaling parameter by a harmonic component.

It Would Not be Obvious to Replace the Weights in Rao Gadde

It would not be obvious to replace the weights W in Rao Gadde with the ratio in Rezayee as suggested by the Examiner. As noted in column 4, lines 21-34 of Rao Gadde, the weights W that are used in Equation 1 are selected based on which weights provide the best speech recognition for a given

signal-to-noise ratio during training. It would not be obvious to arbitrarily replace these carefully chosen weights with the ratio shown in Rezayee, especially when neither Rezayee nor Rao Gadde suggests such a replacement.

In the Office Action, it was asserted that those skilled in the art would have been motivated to replace the weights in Rao Gadde with the ratio in Rezayee for "the improvement in the quality of the voice signal" discussed in paragraph [0021] of Gao. But there is no indication in any of the references that arbitrarily replacing the weights in Rao Gadde with the ratio in Rezayee would produce a better quality voice signal. In addition, even attempting to form a better quality voice signal is counter to the teachings in Rao Gadde. The intent of Rao Gadde is not to form a better quality voice signal but to form a noisier model that better reflects the actual noisy signals that are received. Thus, there would be no motivation to replace the weights in Rao Gadde to provide better quality voice signals, since Rao Gadde is not trying to provide better quality voice signals. Further, it defies common sense to replace weights that have been carefully chosen by Rao Gadde to form noisier models with a ratio that is applied to noisy speech signals to form noise reduced speech signals. As such, it would not be obvious to replace the weights in Rao Gadde as suggested by the Examiner.

The Ratio in Rezayee is Not a Ratio of an Energy of a Harmonic Component in a Frame Without the Random Component of the frame to an Energy of the Frame of the Noisy Speech Signal.

In the Final Office Action, the equation for $g_i(n)$ of Table III of Rezayee was asserted to show a ratio of an energy of a harmonic component of a frame without a random component of the frame to an energy of a noisy speech signal of the frame. However, the computation of $g_i(n)$ does not show such a ratio because the numerator $\lambda_X^i(n)$ in the computation of $g_i(n)$ does not show an energy of a harmonic component for the frame without the random component of the frame.

In Rezayee, the value $\lambda_X^i(n)$ used in the computation of $g_i(n)$ is selected by taking the larger of zero and the difference $(d_i(n) - \lambda_N^i(n))$ between a noisy signal energy $d_i(n)$ of the current frame and noise energy $\lambda_N^i(n)$ computed from a previous frame. Neither the value of zero nor the difference reflects the harmonic component of the current frame.

The difference does not represent the harmonic component of the current frame because it is based on noise from a previous frame and Rezayee indicates that noise changes from frame to frame. ("Noise process is not stationary in general. Of course, the noise process is not assumed stationary, but the practical results show that statistics of most noises gathered from various environments, do not vary rapidly in time.", Rezayee, column 91, left column, last paragraph). Since the noise changes from frame to frame, the difference between a noisy signal energy of a current frame and noise energy computed using noise from a previous frame will not produce the harmonic component of the current frame but instead will produce the harmonic component with some additional residual noise or will produce less than the entire harmonic component. In particular, if the noise in the previous frame is smaller than the noise in the current frame, the subtraction will not remove all of the noise from the noisy speech signal and the result will represent the harmonic component with some residual noise. If the noise in the previous frame is larger than the noise in the current frame, the subtraction will remove part of the harmonic component along with the noise. Note that Rezayee recognizes this latter problem on page 92, left column, last paragraph continuing to the right column where Rezayee indicates that errors in the energy estimation will sometimes cause the difference to be negative, which indicates that the noise estimate from the previous frame may be so different from the noise in the current frame that all of the harmonic component and all of the noise of the current frame is removed by the difference calculation. Thus, since Rezayee indicates that the noise from a previous frame is not the same as the noise in the current frame, the difference used to compute $\lambda_X^i(n)$ does not represent the harmonic component of the current frame.

The value of zero also does not reflect the harmonic component because the computation of the gain $g_i(n)$ is only performed when there is some speech present. (See Rezayee, Table III, where $g_i(n)$ is only computed when T(n)=1 and page 91, last paragraph where T(n)=1 only when speech is detected.) If speech is present, the harmonic component should be greater than zero. However, under Rezayee, the value of $\lambda_X^i(n)$ can be zero even when speech is present. If $\lambda_X^i(n)$ truly represented the harmonic component of speech it would not have a value of zero when speech was present.

Since $\lambda_X^i(n)$ does not represent the harmonic component of a frame, the ratio used to compute $g_i(n)$ in Rezayee does not show the ratio of an energy of a harmonic component in a frame without

the random component of the frame to an energy of the frame of the noisy speech signal as found in claim 1.

II. The Art Does Not Show a Same Fixed Scaling Parameter for All Frames of a Random Component

The combination of cited references does not show a fixed scaling parameter for a random component where the same fixed scaling parameter is used for all frames as found in claim 1.

In the Final Office Action, it was asserted that Gao teaches a scaling parameter Gf that is fixed in paragraphs [0053] and [0054]. However, as noted by Gao, the scaling factor Gf is only fixed when "only background noise (no speech) is detected in the frame." (Gao, paragraph [0053], first sentence) For frames that have speech, the scaling factor produced by Gao changes with each frame as the value of NSR changes with each frame. As a result, the scaling factor in Gao is not the same for all frames as required by claim 1.

Applicants note that as written, claim 1 provides a step of multiplying the random component of a frame by a fixed scaling parameter for the random component for each frame of the noisy speech signal. Thus, although Gao shows a fixed scaling value for every frame that contains nothing but noise, it does not show a fixed scaling parameter that is the same for all frames of a noisy speech signal.

Further, those skilled in the art would not replace the scaling parameter in Gao with a scaling parameter that is fixed so that it is the same for all frames. Under Gao, the gain factor is based on the noise-to-signal ratio NSR. When the noise increases, the gain factor decreases causing a reduction in the signal that is sent between the encoder and decoder. When the noise decreases, the gain factor increases. If the gain factor where fixed for all frames, there would be no need for the gain factor. The purpose of the gain factor is to remove more noise by attenuating an unquantized long-term predictor gain and an unquantized fixed codebook gain when noise is present in the signal. If a fixed value is used, the same gain would be applied to noisy and noise-free signals. This would have the effect of simply reducing the energy of all signals whether they included speech or noise. Those skilled in the art would not perform such an action since it would make it more difficult to hear the speech signal without removing any noise from the speech signal.

None of the other cited references show a fixed scaling parameter of a random component for

all frames of a noisy signal either. As such, the combination of cited references does not show or suggest a fixed scaling parameter for a random component as found in claim 1.

III. Claim 1 Conclusion

Since the prior art does not show a step of determining a scaling parameter for a harmonic component for a frame based on the energy of the harmonic component in the frame and the energy of the noisy speech signal in the frame or using a scaling parameter for a harmonic component that is determined for each frame while using a fixed scaling parameter for a random component for all frames, the combination of cited art does not show or suggest the invention of claim 1 or claims 2, 3, and 5-12, which depend therefrom.

<u>Claim 5</u>

Claim 5 depends from claim 1 and includes further limitations for determining the ratio for the scaling parameter of the harmonic component that make claim 5 additionally patentable over the cited art. In particular, under claim 5, the ratio is formed by summing the energy of samples of the harmonic component, summing the energy of samples of the noisy speech signal and dividing the sum for the harmonic component by the sum for the noisy speech signal.

In the Final Office Action, it was asserted that although none of the references show summing the energy of samples of the harmonic component and the energy of samples of the noisy speech signal, it would be obvious to one of ordinary skill in the art to sum up the energy values of the samples used in Table III of Rezayee to compute the gain $g_i(n)$ to get a more robust estimate for the gain for a frame of a signal or a specific window of time. Applicants respectfully dispute this assertion.

In Table III of Rezayee, a separate gain $g_i(n)$ is determined for each eigenvector i. These gains are placed in a diagonal matrix ($G(n) = diag(g_1(n), g_2(n) \ldots g_K(n))$), which is multiplied by eigenvector matrices. ($\hat{X}(n) = U(n)G(n)U^T(n)Y(n)$ In order for the matrix multiplication to work, the gains for each separate eigenvector must be provided separately. Each gain in turn is determined based on energies determine along the eigenvectors. (Reyazee, page 92, left column, last

paragraph). It is not clear that this determination of energies along an eigenvector can be replaced with a sum of sample energies as suggested by the Examiner and there is clearly no suggestion for replacing the eigenvector energies with such a sum in Reyazee. Further, it is not clear that a sum of samples would work better or at all in place of the eigenvector energies. Without at least some suggestion in the art, it would not occur to those skilled in the art to replace the determination of energies along an eigenvector with a sum of energy samples as suggested by the Examiner. As such, claim 5 is additionally patentable over the cited art.

Claims 13, 15, 16 and 24

Independent claim 13 provides a computer-readable storage medium having computer-executable instructions for performing a series of steps. Those steps include identifying a harmonic component and a random component in a noisy speech signal, wherein identifying the harmonic component comprises modeling the harmonic component as a sum of harmonic sinusoids, each sinusoid having an amplitude parameter. A weighted sum is formed to produce a noise-reduced value representing a noise-reduced speech signal that has reduced noise compared to the noisy speech signal. The weighted sum is formed by multiplying the harmonic component by a scaling value for the harmonic component to form a scaled harmonic component, multiplying the random component by a scaling value for the random component to form a scaled random component, and adding the scaled harmonic component to the scaled random component to form the noise reduced value. The scaling value for the harmonic component is different than the scaling value for the random component. In addition, the scaling value for the harmonic component is separately determined for each frame of the noisy speech signal and the scaling value for the random component is fixed for all frames of the noisy speech signal so that the same scaling parameter for the random component is used at each frame of the noisy speech signal.

Claim 13 is not shown or suggested in the combination of cited art. In particular, none of the cited references show or suggest a scaling value for a harmonic component that is separately determined for each frame of a noisy speech signal together with a scaling value for a random component that is fixed for all frames of the noisy speech signal.

In the Office Action, paragraphs [0053] and [0054] of Gao were cited as showing a scaling

parameter for a random component that is the same for all frames of a noisy speech signal. As indicated above, Gao indicates that the scaling factor will only be the same for frames containing only noise. In other frames of a noisy speech signal, the scaling parameter will changed based on the noise-to-signal ratio. Thus, Gao does not show or suggest a scaling value for a random component that is fixed for all frames of a noisy speech signal.

Since the combination of cited art does not show or suggest the combination of a scaling value for a harmonic component that is separately determined for each frame of a noisy speech signal with a scaling value for a random component that is fixed for all frames of the noisy speech signal, the combination does not show or suggest the invention of claim 13 or claims 15-17, 20, 24, and 25, which depend therefrom.

Claim 20

Claim 20 depends from claim 13 and includes a further limitation wherein determining a scaling value for the harmonic component comprises determining a ratio of an energy of the harmonic component to an energy of the noisy signal.

Claim 20 is not shown or suggested in the combination of cited art. In particular, none of the references shows or suggests determining a scaling parameter for each frame of a harmonic component by determining the ratio of an energy of the harmonic component to an energy of the noisy speech signal.

In the Final Office Action, Fig. 4 and lines 15-20 of Rao Gadde were cited as showing determining a scaling parameter for a harmonic component that is multiplied by the harmonic component and Table III of Rezayee was cited as showing a scaling parameter that is a ratio of an energy of a harmonic component to an energy of a noisy speech signal. Further, it was asserted that it would be obvious to replace the scaling parameter in Rao Gadde with the ratio in Rezayee. Applicants respectfully dispute that it would be obvious to replace the weights shown in Rao Gadde with the ratio shown in Table III of Rezayee and respectfully dispute that the ratio in Rezayee is a ratio of an energy of a harmonic component to an energy of a noisy speech signal.

It Would Not be Obvious to Replace the Weights in Rao Gadde

It would not be obvious to replace the weights W in Rao Gadde with the ratio in Rezayee as

suggested by the Examiner. As noted in column 4, lines 21-34 of Rao Gadde, the weights W that are used in Equation 1 are selected based on which weights provide the best speech recognition for a given signal-to-noise ratio during training. It would not be obvious to arbitrarily replace these carefully chosen weights with the ratio shown in Rezayee, especially when neither Rezayee nor Rao Gadde suggests such a replacement.

In the Office Action, it was asserted that those skilled in the art would have been motivated to replace the weights in Rao Gadde with the ratio in Rezayee for "the improvement in the quality of the voice signal" discussed in paragraph [0021] of Gao. But there is no indication in any of the references that arbitrarily replacing the weights in Rao Gadde with the ratio in Rezayee would produce a better quality voice signal. In addition, even attempting to form a better quality voice signal is counter to the teachings in Rao Gadde. The intent of Rao Gadde is not to form a better quality voice signal but to form a noisier model that better reflects the actual noisy signals that are received. (Rao Gadde, column 3, lines 60-65) Thus, there would be no motivation to replace the weights in Rao Gadde to provide better quality voice signals, since Rao Gadde is not trying to provide better quality voice signals. As such, it would not be obvious to replace the weights in Rao Gadde as suggested by the Examiner.

The Ratio in Rezayee is Not a Ratio of an Energy of a
Harmonic Component to an Energy of a Noisy Speech Signal.

In the Final Office Action, the equation for $g_i(n)$ of Table III of Rezayee was asserted to show a ratio of an energy of a harmonic component to an energy of a noisy speech signal. However, the computation of $g_i(n)$ does not show such a ratio because the numerator $\lambda_X^i(n)$ in the computation of $g_i(n)$ does not show an energy of a harmonic component.

In Rezayee, the value $\lambda_X^i(n)$ used in the computation of $g_i(n)$ is selected by taking the larger of a value of zero and the difference $(d_i(n) - \lambda_N^i(n))$ between a noisy signal energy $d_i(n)$ of the current frame and noise energy $\lambda_N^i(n)$ computed from a previous frame. Neither the difference nor the value of zero reflects a harmonic component.

The difference does not represent a harmonic component because it is based on noise from a

previous frame and Rezayee indicates that noise changes from frame to frame. ("Noise process is not stationary in general. Of course, the noise process is not assumed stationary, but the practical results show that statistics of most noises gathered from various environments, do not vary rapidly in time." , Rezayee, column 91, left column, last paragraph). Since the noise changes from frame to frame, the difference between a noisy signal energy of a current frame and noise energy computed using noise from a previous frame will not produce the harmonic component of the current frame but instead will produce the harmonic component with some additional residual noise or will produce less than the entire harmonic component. In particular, if the noise in the previous frame is smaller than the noise in the current frame, the subtraction will not remove all of the noise from the noisy speech signal and the result will represent the harmonic component with some residual noise. If the noise in the previous frame is larger than the noise in the current frame, the subtraction will remove part of the harmonic component along with the noise. Note that Rezayee recognizes this latter problem on page 92, left column, last paragraph continuing to the right column where Rezayee indicates that errors in the energy estimation will sometimes cause the difference to be negative, which indicates that the noise estimate from the previous frame may be so different from the noise in the current frame that all of the harmonic component and all of the noise of the current frame is removed by the difference calculation. Thus, since Rezayee indicates that the noise from a previous frame is not the same as the noise in the current frame, the difference used to compute $\lambda_X^i(n)$ does not represent a harmonic component.

The value of zero also does not reflect the harmonic component because the computation of the gain $g_i(n)$ is only performed when there is some speech present. (See Rezayee, Table III, where $g_i(n)$ is only computed when T(n)=1 and page 91, last paragraph where T(n)=1 only when speech is detected.) If speech is present, the harmonic component should be greater than zero. However, under Rezayee, the value of $\lambda_X^i(n)$ can be zero even when speech is present. Since a value of zero cannot represent the harmonic component when speech is present, $\lambda_X^i(n)$ does not represent a harmonic component.

Since $\lambda_X^i(n)$ does not represent a harmonic component, the ratio used to compute $g_i(n)$ in the combination of Rezayee with the other cited art does not show the ratio of an energy of a harmonic

component to an energy of the noisy speech signal as found in claim 20.

§103(a) Rejection of claims 7-10, 17, and 25

Claims 7-10, 17, and 25 were rejected under 35 U.S.C. § 103(a) as being unpatentable over Laroche in view of Rao Gadde in view of Gao in view of Rezayee and in further view of Seltzer (SPHINX III Signal Processing Front End Specification, CMU Speech Group, 1999).

Claim 7

Claim 7 depends from claim 6 and includes a further limitation wherein decomposing the noisy speech signal comprises decomposing a vector of time samples into a harmonic component vector of time samples and determining a Mel spectrum for the harmonic component vector.

Claim 7 is patentable over the cited art for the same reasons given above for claims 1 and 6. Seltzer does not provide any of the elements of claim 1 that are missing from the combination of Laroche, Rao Gadde, Gao and Rezayee, as discussed above for claim 1. Therefore, claim 7 is patentable over the combination of Laroche, Rao Gadde, Gao, Rezayee and Seltzer.

Claim 8

Claim 8 depends from claim 7 and includes a further limitation wherein multiplying the harmonic component by the scaling parameter comprises multiplying the Mel spectrum for the harmonic component by the scaling parameter.

In the Final Office Action, it was asserted that section 4d of Seltzer shows the calculation of a Mel Spectrum by the harmonic component with the scaling factor pre-multiplied from an input speech signal. It was further asserted that:

> "The multiplication of the scaling factor could have been pre-multiplied as the frequency content of the signal will not change, but rather the amplitude. However, since the scaling factor applies to all frequency components, the scaling factor can be also multiplied after the Mel Spectrum is obtained, which will allow the same result to be obtained. The multiplying of the scaling parameter to the harmonics or the multiplication of the scaling parameter after the Mel coefficients is found from the harmonics is equivalent. The scaling parameter is the A_k value shown in claim 1. Further, the motivation to have

combined the two references was stated in the rejection for claim 7 for recognizing the synthesized speech created from claim 1."

In Seltzer, section 4d discusses the formation of a Mel Spectrum $\widetilde{S}[l]$. The Mel Spectrum is defined as multiplying a spectrum amplitude at a frequency [k] by a frequency-dependent weight $M_l[k]$, which represents a triangular bandpass filter. Thus, different frequency components are multiplied by different weight values. In Laroche, a deterministic component is defined as a sum of harmonically related sinusoids in equation 1. Each sinusoid has a frequency-dependent amplitude $A_k(t)$, where k indicates the frequency of the sinusoid relative to a fundamental frequency $w_0(t)$.

It appears that the Examiner is asserting that multiplying a Mel Spectrum by a scaling parameter is equivalent to determining a Mel Spectrum of the deterministic component formed by Equation 1 of Laroche because the frequency-dependent amplitudes $A_k(t)$ could be factored out of Equation 1 of Laroche and later multiplied by the Mel Spectrum instead of performing the multiplication within the summand of Equation 1. Applicants respectfully dispute this assertion.

First, the deterministic component of Laroche is defined as the sum of sinusoids of different frequencies, each frequency having a different amplitude $A_k(t)$. As such, the amplitudes $A_k(t)$ cannot be factored out of the summation of equation 1 of Laroche since there is a different amplitude for each sinusoid. In terms of an equation:

$$\sum_{k=1}^{K(t)} A_k(t)\exp(jk(t-t_i)w_0(t)) \neq A_k(t)\sum_{k=1}^{K(t)} \exp(jk(t-t_i)w_0(t))$$

because there is a separate $A_k(t)$ for each frequency component k. If this amplitude cannot be removed from within the summand of Equation 1 of Laroche, there is no way for the multiplication of $A_k(t)$ in equation 1 to be equivalent to multiplying $A_k(t)$ by the Mel Spectrum after forming a Mel Spectrum using Seltzer.

Second, each frequency component S[k] is multiplied by a frequency-dependent weight $M_l[k]$ in Seltzer to form the Mel Spectrum. As such, the frequency-dependent amplitudes $A_k(t)$ cannot be factored out of the summation of the equation of section 4d because each frequency-

dependent amplitude would have to be multiplied by its corresponding frequency-dependent weight $M_l[k]$.

Thus, the Examiner's assertion that multiplying a sinusoid by a frequency-dependent amplitude, taking the sum and then determining a Mel Spectrum is the same as determining a Mel Spectrum of sinusoids without the frequency-dependent amplitudes and then multiplying the resulting Mel Spectrum by the frequency-dependent amplitudes is false. As such, claim 8 is patentable over the cited combination of art.

Claims 9 and 10

Claim 9 depends from claim 8 and includes a further limitation of forming a Mel Frequency Cepstral Coefficients feature vector from the noise-reduced value.

In the Final Office Action, it was asserted that Seltzer taught the forming of a Mel Frequency Cepstral Coefficients feature vector from a speech signal. However, Seltzer does not show or suggest forming a Mel Frequency Cepstral Coefficients feature vector from a noise-reduced value. Instead, Seltzer shows forming the Mel Frequency Cepstral Coefficients vector from an input speech signal.

Further, Seltzer does not show forming such a vector from a noise-reduced value formed by applying a scaling parameter to a Mel Spectrum of a harmonic component as found in claim 9 through its dependency on claim 8.

Similarly, none of the other cited references shows or suggests forming a Mel Frequency Cepstral Coefficients feature vector from a noise-reduced value. As such, claims 9 and 10 are patentable over the combination of cited references.

Claim 17

Claim 17 depends from claim 16 and includes a further limitation wherein identifying a harmonic component comprises identifying a vector of time samples representing a harmonic component and converting the vector of time samples into a Mel spectrum for the harmonic component.

Claim 17 is patentable over the cited art for the same reasons given above for claims 13 and 16. Seltzer does not provide any of the elements of claim 13 that are missing from the

combination of Laroche, Rao Gadde, Gao and Rezayee, as discussed above for claim 13. Therefore, claim 17 is patentable over the combination of Laroche, Rao Gadde, Gao, Rezayee and Seltzer.

Claim 25

Claim 25 depends from claim 24 and includes a further limitation of converting the noise-reduced value into a Mel Frequency Cepstral Coefficients feature vector.

In the Final Office Action, it was asserted that Seltzer taught the forming of a Mel Frequency Cepstral Coefficients feature vector from a speech signal. However, Seltzer does not show or suggest converting a noise-reduced value into a Mel Frequency Cepstral Coefficients feature vector. Instead, Seltzer shows forming the Mel Frequency Cepstral Coefficients vector from an input speech signal.

Similarly, none of the other cited references shows or suggests forming a Mel Frequency Cepstral Coefficients feature vector from a noise-reduced value. As such, claim 25 is patentable over the combination of cited references.

Conclusion

In light of the arguments above, Appellants request that the Board reverse the Examiner's rejection of claims 1-3, 5-13, 15-17, 20, 24, 25.

Respectfully submitted,

WESTMAN, CHAMPLIN & KELLY, P.A.

By:_____
Theodore M. Magee, Reg. No. 39,758
Suite 1400, 900 Second Avenue South
Minneapolis, Minnesota 55402-3319
Phone: (612) 334-3222 Fax:(612) 334-3312

TMM

Claims Appendix

1.   A method of identifying an estimate for a noise-reduced value representing a portion of a noise-reduced speech signal, the method comprising:

decomposing each frame of a noisy speech signal into a harmonic component for the frame and a random component for the frame;

for each frame, determining a separate scaling parameter for the frame for at least the harmonic component wherein determining a scaling parameter for each frame of the harmonic component comprises determining a ratio of an energy of the harmonic component in the frame without the random component of the frame to an energy of the frame of the noisy speech signal;

for each frame, multiplying the harmonic component of the frame by the scaling parameter of the frame for the harmonic component to form a scaled harmonic component for the frame;

for each frame, multiplying the random component of the frame by a fixed scaling parameter for the random component, wherein the fixed scaling parameter is the same for all frames and is less than one to form a scaled random component for the frame; and

for each frame, summing the scaled harmonic component for the frame and the scaled random component for the frame to form the noise-reduced value representing a frame of a noise-reduced speech signal wherein the frame of the noise-reduced speech signal has reduced noise relative to the frame of the noisy speech signal.

2.   The method of claim 1 wherein decomposing the portion of the noisy speech signal comprises modeling the harmonic component as a sum of harmonic sinusoids.

3.   The method of claim 2 wherein decomposing the portion of the noisy speech signal further comprises determining a least-squares solution to identify the harmonic component.

4. (Canceled)

5.   The method of claim 1 wherein determining a ratio comprises:
          summing the energy of samples of the harmonic component;
          summing the energy of samples of the noisy speech signal; and
          dividing the sum for the harmonic component by the sum for the noisy speech
                signal.

6.   The method of claim 1 wherein decomposing the portion of the noisy speech signal comprises decomposing a vector of time samples from a frame of the noisy speech signal into a harmonic component vector of time samples and a random component vector of time samples.

7.   The method of claim 6 further comprising determining a Mel spectrum for the harmonic component from the harmonic component vector of time samples.

8.   The method of claim 7 wherein multiplying the harmonic component by the scaling parameter comprises multiplying the Mel spectrum for the harmonic component by the scaling parameter.

9.   The method of claim 8 further comprising forming a Mel Frequency Cepstral Coefficients feature vector from the noise-reduced value.

10.   The method of claim 9 further comprising using the Mel Frequency Cepstral Coefficients feature vector to perform speech recognition.

11.   The method of claim 1 further comprising using the noise-reduced value to perform speech recognition.

12.   The method of claim 1 further comprising using the noise-reduced value in speech coding.

13.   A computer-readable storage medium having computer-executable instructions for performing steps comprising:

        identifying a harmonic component and a random component in a noisy speech signal wherein identifying the harmonic component comprises modeling the harmonic component as a sum of harmonic sinusoids, each sinusoid having an amplitude parameter;

        forming a weighted sum to produce a noise-reduced value representing a noise-reduced speech signal that has reduced noise compared to the noisy speech signal, wherein the weighted sum is formed by multiplying the harmonic component by a scaling value for the harmonic component to form a scaled harmonic component, multiplying the random component by a scaling value for the random component to form a scaled random component and adding the scaled harmonic component to the scaled random component to produce the noise reduced value, wherein the scaling value for the harmonic component is different than the scaling value for the random component, the scaling value for the harmonic component is separately determined for each frame of the noisy speech signal and the scaling value for the random component is fixed for all frames of the noisy speech signal so that the same scaling parameter for

the random component is used on each frame of the noisy speech signal; and

using the noise-reduced value to perform speech recognition.

14. (Canceled)

15.   The computer-readable storage medium of claim 13 wherein identifying the harmonic component further comprises identifying a least-squares solution.

16.   The computer-readable storage medium of claim 13 wherein identifying the harmonic component comprises identifying a vector of time samples representing a harmonic component.

17.   The computer-readable storage medium of claim 16 wherein identifying the harmonic component further comprises converting the vector of time samples into a Mel spectrum for the harmonic component.

18. (Canceled)

19. (Canceled)

20.   The computer-readable storage medium of claim 13 further comprising determining the scaling value for the harmonic component by determining a ratio of an energy of the harmonic component to an energy of the noisy speech signal.

21. (Canceled)

22. (Canceled)

23. (Canceled)

24.   The computer-readable storage medium of claim 13 wherein using the noise-reduced value to perform speech recognition comprises converting the noise-reduced value into a feature vector and using the feature vector as input to a speech recognition system.

25.   The computer-readable storage medium of claim 24 wherein the feature vector comprises a Mel Frequency Cepstral Coefficient feature vector.

Appendix B - EVIDENCE APPENDIX

(No Evidence)

Appendix C - RELATED PROCEEDINGS

(No Related Proceedings)